编码：项目级-数据库设计说明-Version\_1.0.0-signed-20200828

文档等级：□发布级 ■项目级 □团队级

采纳标准：团队文档质量标准

网络文学大数据分析系统

**Big Data Analysis System of Network Literature**

数据库设计说明

**Version 1.0.0-signed**

编写人：阳玉洁

西工大-华迪实训第四小组

**All Rights Reserved**

**目 录**

**[1 引言 3](#_Toc23960)**

[1.1 标识 3](#_Toc27621)

[1.2 系统概述 3](#_Toc9950)

[1.3 文档概述 3](#_Toc2115)

[1.4 基线 4](#_Toc8535)

**[2 引用文件 4](#_Toc453)**

**[3 外部设计 4](#_Toc29484)**

[3.1 标识符和状态 4](#_Toc26063)

[3.2 使用它的程序 4](#_Toc31890)

[3.3 约定和支持软件 4](#_Toc25272)

[3.4 类型划分 5](#_Toc3495)

[3.5 数据库表设计格式说明 5](#_Toc7844)

**[4 结构设计 6](#_Toc30846)**

[4.1 概念结构设计 6](#_Toc15856)

[4.2 逻辑结构设计 6](#_Toc18504)

[4.3 物理结构设计 7](#_Toc24106)

[4.3.1 数据表设计 7](#_Toc28405)

[4.3.2 数据存取方面的设计 7](#_Toc1228)

[4.3.3 数据存取方面的设计 7](#_Toc2422)

**[5 运用设计 8](#_Toc20934)**

[5.1 数据字典设计 8](#_Toc17758)

[5.2 数据结构设计 10](#_Toc22189)

[5.3 数据流图设计 11](#_Toc10606)

[5.4 安全保密设计 11](#_Toc31824)

# 引言

## 标识

适用系统：网络文学大数据分析系统

中文名称：数据库设计说明（DDS）

英文名称：Database Design Specification (DDS)

文档版本：1.0.0

文档编号：项目级-数据库设计说明-Version\_1.0.0-signed-20200828

## 系统概述

本文档适用于“网络文学大数据分析系统”项目的开发过程。网络文学大数据分析系统项目由西北工业大学软件学院提出，由西工大-华迪实训第四小组成员负责实施，该项目编号为项目级-数据库设计说明-Version\_1.0.0-signed-20200828，其软件产品版本号为“1.0”。

该项目基于B/S（Brower/Server）浏览器/服务器结构，采用传统的三层结构方式进行解耦开发。对当下最热门的网络文学阅读机构——“晋江文学”提供的各项数据进行可视化展示和分析。数据分析采用Hadoop和Spark。

## 文档概述

本文档依据《GB/T 8567-2006 计算机软件文档编制规范》制定，属于技术文档，仅限于项目相关人员阅读。

本文档依赖于软件需求规格说明书，从各方面进行系统的数据库设计规划，用它指导该系统在数据库各方面的内容，为后续项目的开发与测试提供依据。

本文档的保密范围如下：

①开发人员

②测试阶段人员

③对本文档进行评审的人员或机构

④项目组以及其他有权需要调用本文档的人员

## 基线

基线：GB/T 8567-2006

# 引用文件

《GB/T 8567-2006 计算机软件文档编制规范》，国家标准

《项目级-文档编写格式规范-Version\_1.0.0-signed-20200828》，工作室编写

《项目级-软件需求规格说明书-Version\_2.0.0-signed-20200828》，工作室编写

# 外部设计

## 标识符和状态

在我们开发的网络文学大数据分析项目中，总共有3个数据表，它们分别是：书籍信息表，作者信息表和读者打赏信息表。在六个表中均以表格名\_id为主码。

**3.2 使用它的程序**

网络文学大数据分析系统。

**3.3 约定和支持软件**

本系统的开发环境为：

数据库：MySQL 6.5、Redis 5.0.9

操作系统：Windows XP/7/8/9/10、Mac OS X、Linux

开发环境：Spark、Hadoop

开发工具：JetBrains PyCharm 2020

本项目用到的数据项：在名称，范围，类型等方面的约定见数据字典

**3.4 类型划分**

1. 题材收录占比
2. 月份排行
3. 季度排行
4. 年度排行
5. 用户阅读稳定性分析
6. 积分排行
7. 各题材top10
8. 长篇排行
9. 作者积分排行
10. 新晋作者排行
11. 稳定更新排行
12. 本月完结收益排行
13. 打赏收益排行
14. 读者消费排行
15. 读者订阅偏好
16. 同类网站排行

**3.5 数据库表设计格式说明**

表1 数据库表设计说明

|  |  |  |  |
| --- | --- | --- | --- |
| 属性名 | 类型 | 长度 | 编码 |
| 团队内统一定义 | Int/boolean/  verchar/String/Long | 视具体属性而定 | 默认为utf-8 |

# 结构设计

## 概念结构设计

清楚正确地表述本数据库反映的数据形式和联系。管理员可以更新数据，用户仅能浏览数据。

## 逻辑结构设计

**4.2.1 数据库设计规范**

数据库命名规则：项目名称英文全拼，每个英文单词第一个字母大写

表命名规则：均使用表内容对象英语全拼，第一个字母大写

字段命名规则：每个英文单词第一个字母大写

字段时间格式：所有时间格式采用2020-06-25 的形式

**4.2.2 数据库逻辑设计**

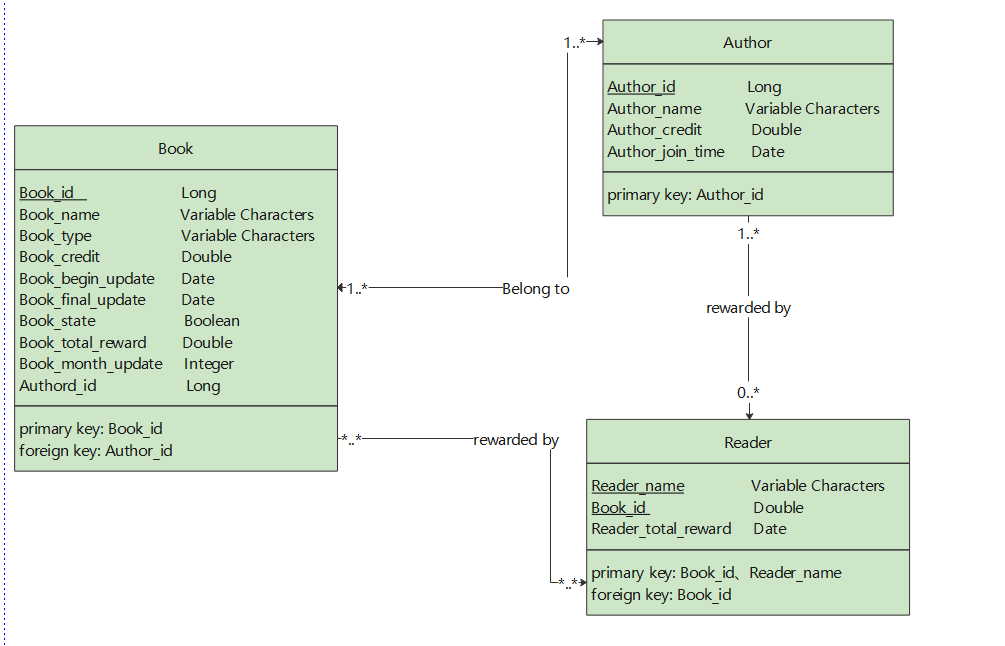


图1 项目整体ER图

## 物理结构设计

### 数据表设计

### 参看数据字典。

### 数据存取方面的设计

对经常在查询中出现的关系的码建立索引；

对经常进行连接操作的关系的码建立索引；

对于更新频率很高的关系模型,没有定义索引，由于技术不成熟，在此不讨论存储位置的设计。

### 数据存取方面的设计

1. 主键约束、非空值约束：

在除读者打赏信息外所有的表格中均用表格对象名\_id为主键。

1. 外键约束：

表2 外键表格

|  |  |
| --- | --- |
| 表格名 | 外键 |
| 读者打赏信息 | 书籍id |
| 书籍信息 | 作者id |
| 删除待保留 | 读者id |

1. 唯一约束：

每个表的主键需定义为唯一性，例如对于作者，作者的编号、作者名称都是用于识别作者身份，所以需建立唯一约束。

1. 定义规则：

参见数据字典。

# 运用设计

## 数据字典设计

读者打赏 = 读者名称 + 书籍id + 打赏数额

作者 = 作者id + 作者名称 + 作者积分 + 作者入站时间

书籍 = 书籍id + 书籍名称 + 书籍题材 + 书籍积分 + 书籍起始更新时 + 书籍最后更新时间 + 状态（完结/连载） + 书籍打赏数额 + 一个月内更新天数 + 作者id

1. 书籍（Book）：根据功能点需求，书籍信息表除了需要书籍id、书籍名称两个标识书籍基本信息的属性外，还需要书籍题材、书籍积分、书籍起始更新时间、书籍最后更新时间、状态（完结/连载）、书籍打赏数额以及一个月内更新天数、作者id这些信息。从功能上数据表设计如表3所示。

表3 书籍信息表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 列名 | 数据类型 | 长度 | 说明 | 备注 |
| Book\_id | Long | 8 | 书籍id | 主键 |
| Book\_name | Varchar | 50 | 书籍名称 |  |
| Book\_type | Varchar | 20 | 书籍题材 |  |
| Book\_credit | double | 15 | 书籍积分 |  |
| Book\_begin\_update | date | 20 | 书籍起始更新时间 |  |
| Book\_final\_update | date | 20 | 书籍最后更新时间 |  |
| Book\_state | Boolean | 4 | 状态（完结/连载）  完结状态为0，连载状态为1 |  |
| Book\_total\_reward | double | 15 | 书籍打赏数额 |  |
| Book\_month\_update | Int | 4 | 一个月内更新天数 |  |
| Author\_id | Long | 8 | 作者id | 外键 |

1. 作者（Author）：根据功能点需求，作者信息表除了需要作者id、作者名称两个标识作者基本信息的属性外，还需要作者积分和作者入站时间两个信息。从功能上数据表设计如表4所示。

表4 作者信息表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 列名 | 数据类型 | 长度 | 说明 | 备注 |
| Author\_id | Long | 8 | 作者id | 主键 |
| Author\_name | Varchar | 50 | 作者名称 |  |
| Author\_credit | double | 15 | 作者积分 |  |
| Author\_join\_time | date | 20 | 作者入站时间 |  |

1. 读者（Reader）：根据功能点需求，读者打赏信息表除了需要读者名称标识读者基本信息的属性外，还需要书籍id和读者打赏数额两个信息。从功能上数据表设计如表5所示。

表5 读者打赏信息表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 列名 | 数据类型 | 长度 | 说明 | 备注 |
| Reader\_name | Varchar | 50 | 读者名称 | 与Book\_id一起作为复合主键 |
| Book\_id | Long | 8 | 书籍id | 外键 |
| Reader\_total\_reward | double | 15 | 读者打赏数额 |  |

1. 外部实体

表6 外部实体

|  |  |  |  |
| --- | --- | --- | --- |
| 名称 | 说明 | 输入数据流 | 输出数据流 |
| 管理员 | 后台管理的主要实体 | 新的数据源 | 各种数据更新 |

## 数据结构设计

表7 数据结构设计

|  |  |
| --- | --- |
| 数据结构名 | 属 性 |
| 书籍信息 | 书籍id、书籍名称、书籍题材、书籍积分、书籍起始更新时间、书籍最后更新时间、状态（完结/连载）、书籍打赏数额、一个月内更新天数、作者id |
| 作者信息 | 作者id、作者名称、作者积分、作者入站时间 |
| 读者打赏信息 | 读者名称、书籍id、读者打赏数额 |

## 数据流图设计

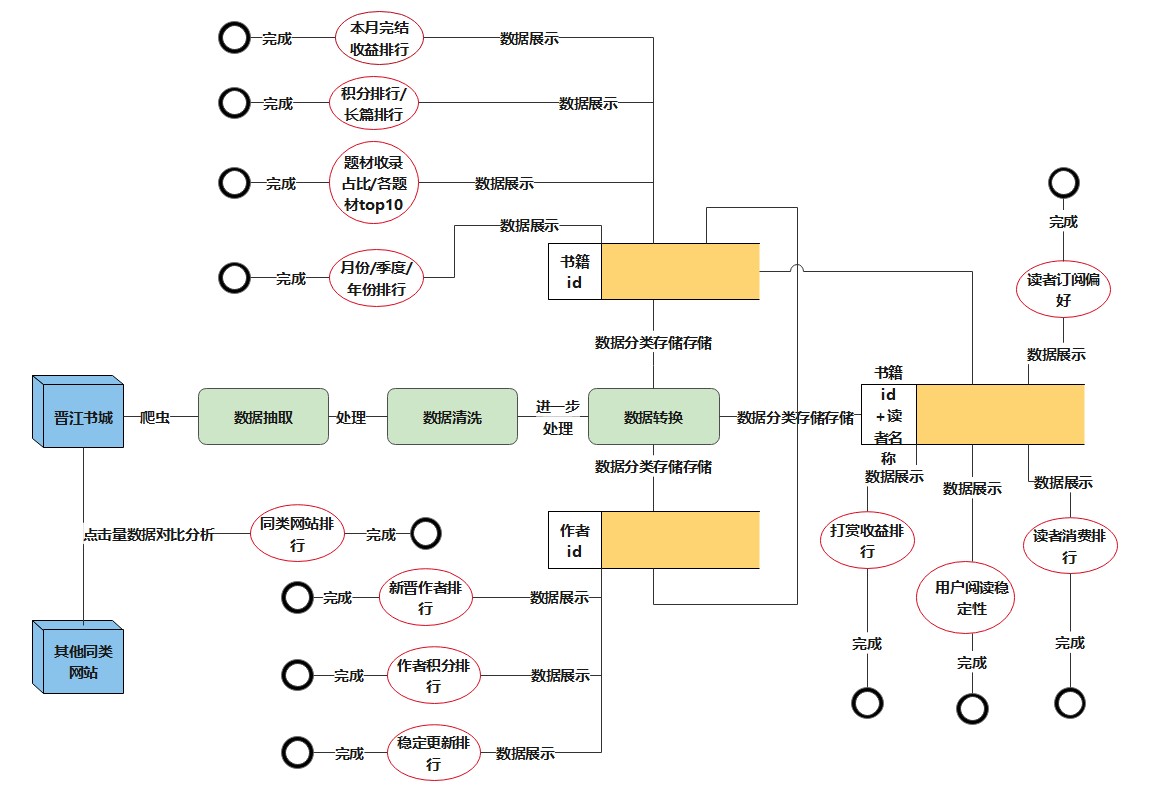
****

图2 数据流图

## 安全保密设计

数据库作为管理系统的基础，通常保存着重要的商品信息和用户信息。由于涉及到财务管理，数据的完整性和安全性显得尤为重要。系统中的数据一旦丢失，将需要很长时间进行恢复，有时甚至使信息系统不得不从系统初始化阶段重新开始运行。每天进行数据备份是保障系统安全的重要手段。数据备份需要严格按照事先制定的备份与故障恢复策略进行，并落实备份登记和检查措施。